THE RESOLUTION GAME
A Multiple Selves Perspective

DIMITRI MIGROW[*]

University of Regensburg, Economics Department

AND MATTHIAS UHL[†]

Max Planck Institute of Economics, Jena

The notion of choice inconsistency is widely spread in the literature on behavioral economics. Several approaches were used to account for the observation that people reverse their choices over time. This paper aims to explain the formation of resolutions regarded as internal self-binding devices. It moves away from anthropocentric neoclassicism and embraces a more atomistic notion of a player by defining intrapersonal agents as strategic actors. The magnitude of state-dependency is seen as a key driver of intrapersonal conflict modelled by the incongruity of the preferences of two opposing agents. The sequential conceptualisation basically allows for experimental testing.

KEYWORDS: Multiple Selves, Agency, Intrapersonal Conflict, Resolutions, Self-Binding, Self-Control, Commitment.

1. INTRODUCTION

We are often not at one with ourselves. The empirical observation that people tend to reverse choices as time passes is not new and has challenged economists for quite some time. One basic explanation is the existence of uncertainty and the fact that people update their beliefs. Harder to match with neoclassical theory is the observation that some people are obviously willing to spend scarce resources to restrict their options in the future. From the viewpoint of neoclassical economics it would be intuitive to keep as many possibilities as you can. At a later point in time you can always choose irrespective of your less preferred options but you have them available for the case of having gained more valuable information. Note that the problem is not simply about reducing one's searching costs, for instance, by asking the waiter to only bring the first page of a menu. One intentionally excludes certain future options from his choice set. This fact denotes that people are aware of systematic inconsistencies in their choice behavior and do not like that. Apart from the psychological appeal of the problem, it is illuminating from a theorist's perspective as it raises methodologically relevant questions. These boil down to the one central question: Whom do we consider as the actor in an economic model?

We aim to explain the formation of resolutions and the emergence of their certain extent. First, we want to take the psychological notion of intrapersonal

[*] dimitri.migrow@uni-regensburg.de

[†] uhl@econ.mpg.de

conflict serious by modelling this behavioral ambivalence. Second, we present a model that creates falsifiable hypotheses and is therefore open for empirical testing of its predictions. For this reason it is restricted to only a few measurable parameters. We use a dual selves model. The motivation for this concept is briefly addressed in the second section. The third section describes the psychological superstructure of the problem we consider here with the help of an example. The fourth and fifth section contain the formal model. The sixth section offers a conclusion.

2. 'MULTIPLE SELVES' AS AGENTS

The notion of multiple selves becomes an interesting tool if we think of a conflict within a person. The word 'conflict' makes it meaningful to regard the person not as a unified decision-maker but as a composition of at least two decision-makers who are engaged in strategic interaction.[1] MOLDOVEANU AND STEVENSON (2001) point out that the so-called Aristotelian modelling tradition is dominant in the social sciences. This view considers the person as a unitary preference system which equates the biological entity with the entity of decision-making. The so-called Heraclitean tradition on the other hand considers persons as fragmented, internally torn apart, "or as an ongoing and irresolvable conflict of competing interests, impulses, or identities." (MOLDOVEANU AND STEVENSON 2001, p. 295) While in game theory it is quite common to model a person as a multitude of decision-makers (see, for instance, GÜTH 1991), the Aristotelian tradition is the orthodox perspective of microeconomists. For some, maybe even for most economic problems this can be a useful assumption.

But there is nothing in the foundations of neoclassical economics that forces us to view the biological person as the primitive actor. In our view, there exists a class of microeconomic decision problems where the biological entity as the primitive actor is not very suitable. In these situations decomposing the person into several agents can be rewarding. It enables us to account for phenomena of behavioral economics without the necessity to forgo the analytical rigour of neoclassical theory. The concept of agency allows us to stick to the convenient assumption of stable preferences as an agent cannot experience a preference reversal by definition. This is because clear intentions due to stable preferences are exactly what constitutes an economic agent. The selves in our model are nothing but economic agents with conflicting preferences. In our opinion, forming resolutions and taking actions against future options is a problem in which multiple selves can add theoretical insights.

Thomas Schelling is the most popular advocate of the notion of multiple selves in economic literature, he argues that "Everybody behaves like two people, one who wants clean lungs and long life and the other who adores tobacco, or one

[1] We will avoid the term 'individual' as from its root it means 'the indivisible'. In our methodological context a person should therefore not be seen as an individual but as a dividual.

who wants a lean body and the other who wants dessert" (Schelling 1984, p. 58). There are several attempts by other authors to formalize Schelling's ideas. Most of these papers focus on the time aspect of choice behavior and emphasize a conflict between short-term and long-term interests (see, for instance, Thaler and Shefrin 1981, as a classical reference, or O'Donoghue and Rabin 2005 and Fudenberg and Levine 2006 for more recent approaches). This is related to hyperbolic discounting models (see, for instance, Strotz 1956 or Laibson 1997) where choice inconsistencies result due to a perspective change as the person advances on a timeline. The persons in these models have an 'immediacy bias' expressed by a hyperbolic discounting function.

Our focus is different. Even though timing plays a role in our analysis it is not the explanatory driver of the intrapersonal conflict we have in mind. We examine the conflict a person experiences due to different preferences in different states of the world abandoning any discounting. We encounter state-dependency if the state of the world is of direct concern to the preference relation of a person (Karni 1993, p. 188). The uncertainty in our model is connected to the upcoming social environment and not concerning the type of one's own future selves which is an alternative way to capture uncertainty. In contrast to hyperbolic discounting models with uncertainty about one's own future type we keep preferences fixed and make them common knowledge. This makes our problem more salient. It guarantees for an inner conflict, whereas a hyperbolic discounting framework with unknown 'immediacy bias' results in a lottery of whether or not being able to stick to one's plans. In such as setting, exposing oneself to certain situations would mean to gamble and make it somehow cumbersome to interpret resolutions. It would then be more meaningful to raise the question whether one enters the risk of getting plans spoiled by a 'weaker self' or whether one chooses some kind of outside option. Additionally, in our view, the focus on time in hyperbolic discounting models makes personal welfare calculations rather difficult, since it is not clear at what point in time one should employ them.

Apart from that most of the papers focus on the use of external commitment devices to account for a lack of willpower. This is despite the fact that personal rules like resolutions are probably more common (Ainslie 2001, pp. 78 - 85). Bénabou and Tirole (2004) is a rare example of analyzing personal rules as an internal commitment device. Embedded in an enriched quasi-hyperbolic discounting framework they are emphasizing the aspect of imperfect recall concerning one's own motives. In their view, when uncertain about my own willpower, my past choices can give an indication of what kind of person I am and give me self-confidence. In the presence of imperfect recall, resolutions can be compared to journal keeping and help to regularly scrutinize one's behavior (Bénabou and Tirole 2004, p. 879). Resolutions activate lapse-related memory ex post or make some actions more salient ex ante. An interesting implication of their model is that over time a subject can employ overly rigid rules or compulsive behavior, if self-signaling becomes an end in itself. This means that people regard any situation as a test of their willpower even if self-restraint is not desirable ex ante (Bénabou and Tirole 2004, p. 851). Their concept of resolutions is more complex than the one discussed here since they are taking a dynamic perspective and focus on memory determination. In our static approach, as opposed to that, a resolution should not be compared to a cognitive rule but to a quantifiable concession combined with a consumption barrier.

This interpretation is closer to Brocas and Carrillo (2008), who propose a neurologically motivated principal-agent-approach. But their focus is clearly on informational asymmetries. They emphasize the aspect of 'incentive salience' which describes a bias in the utility function of the myopic agent towards excessive consumption of an enjoyable good. Under complete information the principal could impose her optimal choices. Under incomplete information the principal is forced to implement a revelation mechanism to induce self-selection of the agent which costs resources. The decision problem arises as a tension between inducing self-selection and managing resources (Brocas and Carrillo 2008, pp. 1330 - 1331). In this view, there exists a multitude of agents, where each time-indexed agent learns about his current willingness to consume once he appears. The principal only knows the distribution of agent types.

Note that the selves in our model are neither suffering from naivity or imperfect recall concerning their future or past choices, nor are they subject to informational asymmetries. 'Incentive salience' in our context results only from diverging evaluations of the same prospect depending on the state a person is in. Therefore we use symmetric utility functions for both selves. Evaluations are stable and common knowledge, thus, we restrict our analysis to two deterministic selves. Uncertainty stems from the fact that one does not know the setting in which a good will be consumed, and accordingly, how costly it will be to break a formed resolution. In our model, there is no principal-agent hierarchy between both selves but a strategic advantage of the planner. The restriction to two selves is of course arbitrary from a psychological or philosophical perspective but due to reasons of simplicity given our explanatory goal. With a dual selves model and our specific uncertainty context quantitative resolutions can be captured and analyzed in a reasonable way.

3. AN EXAMPLE OF INTRAPERSONAL CONFLICT

Consider a person who in the afternoon is invited to an evening party. The person has a split attitude towards alcohol. When being at home she dislikes alcohol because she thinks it is unhealthy. From her view at home the less alcohol, the better it is. But when being at parties she always enjoys drinking. She would then like to drink excessively. She is aware of this behavioral pattern.[2] This

[2] Smith uses the allegory of an impartial spectator who is able to judge from a distance. Now assume that the person at home is aware of this behavioral pattern and wants to prevent it."There are two different occasions when we examine our own conduct, and endeavour to view it in the light in which the impartial spectator would view it; first when we are about to act, and secondly, after we have acted. (...) When we are about to act, the eagerness of passion

fact relates to the idea of 'sophistication' in hyperbolic discounting models (see, for instance, O'Donoghue and Rabin 1999), where it means to be aware of your 'immediacy bias' which will spoil your plans. 'Sophistication' means not to underestimate this 'immediacy bias' even though you might not precisely know its magnitude (O'Donoghue and Rabin 2001). If you are unsure whether this bias exists at all, as it depends on factors you cannot foresee, you are uncertain about the type of your own future selves. Accordingly, you do not know what your later preferences will be.

We are considering 'sophistication' in a different uncertainty context that we find suitable for the analysis of resolution formation. In our analysis the person at home is sure about her inverted preferences at the party. Therefore she knows that she will have to deal with an opponent. Anticipating this tension the person at home can form a resolution. This could be the strictest possible resolution of abstinence or a generous resolution to drink in moderation. We assume that the resolution sets up an inhibition threshold that is psychologically costly to overcome. Economically speaking, such a resolution would be useless if it would be costless to break. The height of this inhibition threshold depends on the social environment at the party. It is clear to the person ex ante that the social environment at a friend's birthday party is more likely to cause a low inhibition threshold than the one at her grandparent's golden wedding, for example.

A second consequence of the resolution is that it sharpens the awareness of consumption ambivalence. If we had not regarded the consumption as problematic why would we have formed a resolution in the first place? If the person at the party adheres to the resolution, this ambivalence dilutes the joy of consumption through a cognitive dissonance (Festinger 1957). Succeeding to say "The heck with it, drinking is what parties are there for!" means achieving cognitive consonance *and* breaking the resolution. The person at the party has liberated herself and is no longer 'locked' in the resolution world of ambivalent consumption, i.e. cognitive dissonance.

Without a resolution there would not be any inhibition threshold or cognitive dissonance and the person at the party would simply get drunk. This prospect is seen as worst case from the person at home. Forming a resolution is therefore a dominant strategy for the person at home. But for her it is often not rational to form the strictest possible resolution of abstinence, which could seem natural at first glance. The person at home is encountering the following dilemma. Given her preferences she prefers abstinence. But she knows that the person at the party would be frustrated and breaking the resolution becomes the more attractive the more frustrated she is. If sufficiently frustrated the person at the party would even at relatively high cost caused by a high inhibition threshold break the resolution. To avoid this worst case, the person at home forms a more generous resolution allowing the person at the party to drink some alcohol to reduce frustration.

will seldom allow us to consider what we are doing with the candour of an indifferent person. (...) The fury of our own passions constantly calls us back to our own place, where every thing appears magnified and misrepresented by self-love." (Smith 1759, pp. 261 - 262)

From this reasoning we can clearly see that a generous resolution of the person at home is not motivated by altruism but by prevention of excess consumption at the party.

Note that this is of course only one example out of a larger class of problems to which our logic applies. Alternatively, think of any state-dependent habit where a person in a complement state forms a resolution to restrict this habit. In this state she prefers herself to abstain from the habit but fears intemperance once her 'weaker self' crosses the frontier. It is likely that a strict resolution will be broken in an environment that causes low inhibition thresholds by trend. A limited concession is then considered as the lesser of two evils. We speak of any situation where preferences are clear but the environment in which a habit in question is indulged is uncertain. Take as another example a person that plans to write a thesis in the afternoon. When planning, she is convinced that it is best to watch no television at all. But she also knows that this is likely to lead to such a level of frustration when the afternoon program starts that she will break the resolution and watch the whole program. Thus, she explicitly allows herself to consume precisely one or precisely two hours of television. The costs of breaking this resolution depend on the quality of the program and she has a probilistic idea about this. As a third example think of a social smoker who is inevitably feeling the urge to smoke in the presence of other smokers. Limitting herself ex ante to abstinence will probably lead to such a frustration at the club that she buys a new packet and consumes it once it is at hand. Again, restricting herself to a specified amount of just a few can be rational from the planning perspective given that there is a chance to meet other smokers at the club. This probability she can judge.

In the following chapters we will explicate these intuitions by analysing the problem again in a formal language. We will set up a model of dual selves and stick to a framework where a person at home plans for some social situation she is going to face. Herein the person at home is considered as the first economic agent who we call the planning self and the person facing the social situation is considered as the second economic agent who we call the enjoying self.

4. MODEL SETUP

The basic logic of our model is inspired by a model of taxation as the one used in ACEMOGLU AND ROBINSON (2001, 2006). Another analogy to their work is our understanding of a breach of resolution which is similar to their interpretation of a revolution. This is not surprising, since in a way, Acemoglu and Robinson are considering the distributional conflict within a society and so are we. While their society is composed of conflicting classes, ours is composed of conflicting selves. As our model functions like a taxation model it requires an interagent comparability of utilities since utils are transferred from one agent to the other during the course of the game. What one agent gets, the other agent has to give up. Relying on a concept of interagent transferability of utilities, our approach has the advantage of being able to discuss personal welfare implications similar to social welfare implications, underlining once more that our person is in fact a (small) society of selves.

Consider two players or agents, $S_i, i = \{p, e\}$, referred to as planning self (S_p) and enjoying self (S_e). Both agents together form the identity of a biological person. Each of them takes control over the body at different points in time. The respective agent can determine the acts of the person alone for the time of his 'reign'. At the beginning of the period, in $t = 0$, S_p decides about the acts of the person.

4.1. *Timing*

The course of a period is as follows:

1. In $t = 0$ S_p decides about its level of concessions, denoted by $\gamma \in [0, 1]$, and accordingly forms a resolution.
2. In $t = 1$ nature determines the realization of the environmental variable, denoted by $\mu \in [0, 1]$.
3. In $t = 2$ S_p is passive but able to remind S_e of the formed resolution. S_e observes the realization of μ and decides whether to adhere to the resolution or whether to break it, denoted by $\rho \in \{0, 1\}$.
4. In $t = 3$ both agents experience their payoffs.

4.2. *The Inner Disbalance as Driver of Conflict*

The utility function of S_i is given by u^i. S_p can perfectly anticipate S_e's utility. By the parameter θ we measure the inner disbalance of the person by comparing the initial utilities of both selves. By initial utilities we mean the utilities of consumption abstinence in the social situation, $u^i(A)$.

$$\theta = \frac{u^p(A)}{u^p(A) + u^e(A)}. \tag{1}$$

We assume that $u^p(A) > u^e(A) > 0$. Thus, it follows that $\theta \in (\frac{1}{2}, 1)$. The crucial aspect we capture here is that S_e values consumption abstinence in a social situation less than S_p. Thus, there exists always an inner disbalance, which is more severe if the divergence between the initial utilities of both selves, and accordingly θ, is larger. Technically, θ is the magnitude of state-dependency since for a relatively high θ the state matters more for the relative judgement of a given prospect. In this case, moving from the state of planning for to the state of acting in a social situation causes a larger divergence between the state-dependent utilities for the prospect 'abstinence in social situation'.

4.3. *Concessions by Generous Resolutions*

S_p as Stackelberg leader can improve S_e's well-being in the social situation by making concessions in form of a generous resolution.

We assume that utilities after granting concessions, γ, are given by

$$u^i(\gamma) = (1-\gamma)u^i(A) + (\gamma - C(\gamma))\frac{u^p(A)+u^e(A)}{2}. \tag{2}$$

Note that higher concessions mean higher consumption of an ambivalent good. The first term of (2) captures the negative concession aspect which is less important for S_e than for S_p. This is true since S_e's initial utility is lower by assumption, $u^p(A) > u^e(A)$. The second term of (2) captures the positive concession aspect which is, vice versa, more important for S_e than for S_p for the same reason. The positive concession aspect is diluted by cognitive dissonance, $C(\gamma)$, where $\gamma \geq C(\gamma)$ $\forall \gamma$. We assume $\frac{dC}{d\gamma} > 0$ and $\frac{d^2C}{d\gamma^2} > 0$, which means cognitive dissonance gets increasingly worse as the level of concessions rises. The larger θ and the lower $C(\gamma)$, the stronger is the redistribution of utils from S_p to S_e for any given level of concessions. Since we assume utility symmetry between the two selves, redistribution effects are solely driven by the divergence of initial utilities. Thus, (2) works like a lump sum transfer in taxation models (see Acemoglu and Robinson 2006, pp. 101 - 103).

4.4. *Breach of Resolution*

The environmental parameter μ is a random variable. Let $f(\mu)$ be the density function of μ, where $\int_{\mu=0}^{1} f(\mu)d\mu = 1$. The environmental parameter determines which fraction of S_e's utility will be lost if it breaks the resolution. This represents the height of S_e's inhibition threshold. The idea is that overcoming one's inhibitions to break the resolution and to achieve cognitive consonance is costly. This is an environmental parameter since the intensity of psychological effort and therefore the height of the inhibition threshold depends on the social environment.

S_p knows the distribution of inhibition thresholds, meaning that it has a realistic expectation about the consumption adversity, i.e. the applied social norm, of the social environment. A social environment with strong social norms against consumption would mean a right skewed distribution of μ. This makes a low inhibition threshold ceteris paribus more likely than a neutral environment. S_p cannot influence the social environment, nor can it avoid that S_e faces it except by staying at home. This is not considered as an alternative in our framework by assumption.

The action variable of S_e concerns the breach of resolution and is expressed by ρ. We assume the following.

If S_e adheres to the resolution, $\rho = 0$, the utility levels, u^i, are given by (2).

If S_e breaks the resolution, $\rho = 1$, the utility levels are given by

$$u^e(\rho = 1|\gamma) = (1 - \mu)(u^p(A) + u^e(A)), \tag{3}$$

and

$$u^p(\rho = 1|\gamma) = 0. \tag{4}$$

Comparing (3) and (4) we can see that by a breach of resolution S_e can only improve its well-being at the expense of S_p. The second bracket of the right-hand side of (3) is due to the fact that S_e annexes all of S_p's utility by a breach of resolution. The first bracket expresses that for doing so it has to cross its inhibition threshold paying a total utility fraction of μ. (4) is the other side of the medal: all of S_p's utility is annexed as the outcome is no longer a compromise. Once S_e has broken the resolution and achieved cognitive consonance it will know no bounds and consume its satiation level. This is worst case from the perspective of S_p causing it a zero utility. Note that this would have to be expected in any case without a resolution as by our definition there would not be any cognitive dissonance and no inhibitions for S_e in the first place. For S_p forming a resolution is a dominant strategy by assumption.

5. MODEL ANALYSIS

Subsection 5.1. determines the emerging extent of resolutions chosen by S_p derived from backward induction. Subsection 5.2. considers the welfare implications of the equilibrium resolution and contrasts them with external self-binding devices.

5.1. *The Optimal Resolution*

Remember that the preferred prospect from the perspective of S_p would be consumption abstinence (A) giving it an unreduced utility of $u^p(A)$. S_e's utility in this case is $u^e(A)$, where $u^p(A) > u^e(A)$.

To obtain the optimal resolution S_p chooses, let us first consider the range of μ at which S_e adheres to the resolution. This range follows from the level of concessions, $\gamma \in [0, 1]$, made by S_p. A resolution is adhered to if $u^e(\rho = 0|\gamma) \geq u^e(\rho = 1|\gamma)$ implying

$$(1 - \gamma)u^e(A) + \frac{(\gamma - C(\gamma))(u^p(A) + u^e(A))}{2} \geq (1 - \mu)(u^p(A) + u^e(A)),$$

or, by exploitation of (1),

$$\mu \geq \theta - \theta\gamma + \frac{\gamma + C(\gamma)}{2}. \tag{5}$$

Thus, we can define a critical value of the environmental parameter, $\widetilde{\mu}$, by

$$\widetilde{\mu} = \theta - \theta\gamma + \frac{\gamma + C(\gamma)}{2}, \tag{6}$$

where S_e is just indifferent between adhering to and breaking the resolution. Remember that the magnitude of the environmental parameter represents the height of the inhibition threshold to break the resolution.

The crucial point is that S_p can influence this critical value via the level of concessions respectively the generosity of the resolution it chooses.

The first and second derivative of (6) are given by

$$\frac{d\widetilde{\mu}}{d\gamma} = -\theta + \frac{1}{2} + \frac{1}{2}\frac{dC(\gamma)}{d\gamma}, \tag{7}$$

and

$$\frac{d^2\widetilde{\mu}}{d\gamma^2} = \frac{1}{2}\frac{d^2C(\gamma)}{d\gamma^2} \tag{8}$$

From (7) we can see that the critical value can be reduced by making higher concessions respectively forming more generous resolutions till the level of concessions equals γ^*, which is implicitly given by $\frac{dC(\gamma^*)}{d\gamma} = 2\theta - 1$. From this level on, making higher concessions increases the critical value again as the convexity of cognitive dissonance becomes rampant. $\widetilde{\mu}$ is therefore U-shaped in γ. This is illustrated by the upper curve in Figure 1, where we assumed $C(\gamma) = \frac{1}{2}\gamma^2$. If μ takes on a value below this curve S_e will break the resolution.

S_p's utility is

$$u^p(\gamma) = \begin{cases} 0, & \text{if } \mu \in [0, \widetilde{\mu}), \\ (1-\gamma)u^p(A) + \frac{\gamma - C(\gamma)}{2}, & \text{if } \mu \in [\widetilde{\mu}, 1]. \end{cases} \tag{9}$$

Given (6) and (9) we can derive an optimal level of concessions made by S_p.

Proposition 1. *A unique optimal level of concessions made by S_p in a resolution game is given by*

$$\frac{f(\widetilde{\mu}(\widehat{\gamma}|\theta))}{1 - F(\widetilde{\mu}(\widehat{\gamma}|\theta))} = \frac{\theta + \frac{\widehat{\gamma}-1}{2}}{(\frac{C(\widehat{\gamma})-\widehat{\gamma}}{2} - (1-\widehat{\gamma})\theta)\frac{d\widetilde{\mu}(\widehat{\gamma}|\theta)}{d\gamma}}. \tag{10}$$

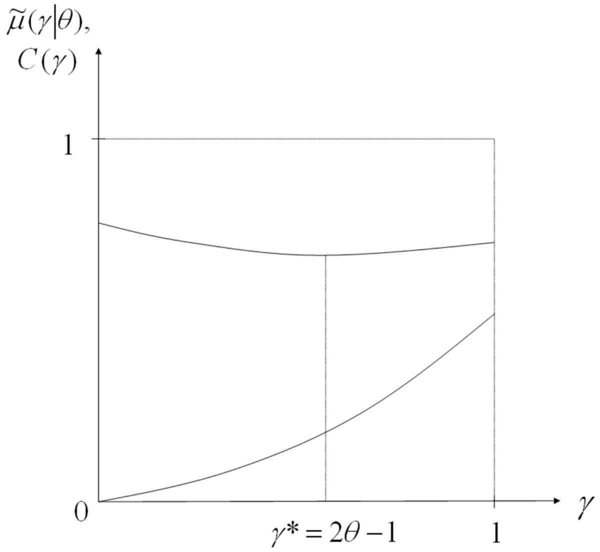

FIGURE 1.— Critical environmental parameter and cognitive dissonance

Proof. See appendix. ■

Proposition 2. *An optimal concession is characterized by* $\widehat{\gamma} < \gamma^*$.

Proof. Given the positive left-hand side of (10), note that the expression on the right-hand side is positive if and only if $\widehat{\gamma} < \gamma^*$. While $(\frac{C(\widehat{\gamma})-\widehat{\gamma}}{2} - (1-\widehat{\gamma})\theta)$ is always negative since $\gamma \geq C(\gamma)\ \forall\gamma$ by assumption, the right-hand side gets positive only in case $\frac{d\widetilde{\mu}(\widehat{\gamma}|\theta)}{d\gamma} < 0$. Given the U-shaped form of $\widetilde{\mu}(\gamma|\theta)$ with $\widetilde{\mu}(\gamma^*|\theta) < \widetilde{\mu}(\gamma'|\theta)\ \forall\gamma' \neq \gamma^*$, clearly $\frac{d\widetilde{\mu}(\widehat{\gamma}|\theta)}{d\gamma} < 0$ only if $\gamma < \gamma^*$. Thus, $\widehat{\gamma} < \gamma^*$. ■

Proposition 3. *With an increasing inner disbalance* S_p*'s optimal concession respectively its optimal resolution increases as well.*

Proof. Since $\frac{\partial F(\widetilde{\mu}(\widehat{\gamma}|\theta))}{\partial\widehat{\gamma}} \neq 0,\ F(\widetilde{\mu}(\widehat{\gamma}|\theta))$ defines $\widehat{\gamma}$ as an implicit function of θ. The question is now, how $\widehat{\gamma}$ changes if θ changes marginally. By total differentiation we get

$$\frac{d\widehat{\gamma}}{d\theta} = -\frac{\partial F(\widetilde{\mu}(\widehat{\gamma}|\theta))/\partial\theta}{\partial F(\widetilde{\mu}(\widehat{\gamma}|\theta))/\partial\widehat{\gamma}},$$

or

$$\frac{\partial\widetilde{\mu}(\widehat{\gamma}|\theta)}{\partial\widehat{\gamma}} = -\frac{\frac{\partial\widetilde{\mu}(\widehat{\gamma}|\theta)}{\partial\theta}}{\frac{d\widehat{\gamma}}{d\theta}}. \tag{11}$$

By substituting (11) into (10) we obtain

$$\frac{f(\widetilde{\mu}(\widehat{\gamma}|\theta))}{1-F(\widetilde{\mu}(\widehat{\gamma}|\theta))} = \frac{(\theta+\frac{\widehat{\gamma}-1}{2})\frac{d\widehat{\gamma}}{d\theta}}{-(\frac{C(\widehat{\gamma})-\widehat{\gamma}}{2}-(1-\widehat{\gamma})\theta)\frac{\partial\widetilde{\mu}(\widehat{\gamma}|\theta)}{\partial\theta}}. \tag{12}$$

Since $\frac{\partial\widetilde{\mu}(\widehat{\gamma}|\theta)}{d\theta} > 0$, (12) shows that $\frac{d\widehat{\gamma}}{d\theta} > 0$. ■

5.2. *Personal Welfare Implications*

Assume the existence of a benevolent dictator who is considering paternalistic intervention for the sake of the person, for example by imposing a prohibition of consumption. The welfare implications of our analysis that we will discuss in the following are considerations of an omniscient outside observer. Since we analyzed the person as a society of selves it seems natural to look at social welfare criteria which for our case will mean personal welfare criteria. Using a Benthamite personal welfare function the utility of the person, u^T, is the sum of the utilities of both agents. If S_e adheres to the resolution u^T is specified by the sum of u^p and u^e being

$$u^T(\rho=0|\gamma) = (1-C(\gamma))(u^p(A)+u^e(A)). \tag{13}$$

If S_e breaks the resolution the person's utility is specified by the sum of (3) and (4) which is, since S_p's utility is reduced to zero, simply S_e's utility and therefore

$$u^T(\rho=1|\gamma) = (1-\mu)(u^p(A)+u^e(A)). \tag{14}$$

Thus, we can see that within a Benthamite framework for any given concessional level it is welfare maximizing from the person's perspective that S_e adheres to the corresponding resolution if $u^T(\rho=0|\gamma) \geq u^T(\rho=1|\gamma)$ implying

$$(1-C(\gamma))(u^p(A)+u^e(A)) \geq (1-\mu)(u^p(A)+u^e(A)),$$

or

$$\mu \geq C(\gamma). \tag{15}$$

If the inhibition threshold to break any given resolution is weakly larger than the cognitive dissonance caused by adhering to this resolution, it is personal welfare maximizing to adhere to it. Vice versa, if the inhibition threshold is lower, it is personal welfare maximizing to break it. Since this is true for any concessional level and its corresponding resolution it is also true, especially, for $\gamma = \widehat{\gamma}$ being S_p's optimal concession. The lower curve in Figure 1 represents the cognitive dissonance. If μ takes on a value below this curve it is welfare optimal to break the resolution.

Proposition 4. *In equilibrium a resolution is broken more often than a welfare maximum within a Benthamite framework suggests.*

Proof. Actually, any given resolution will be adhered to by S_e if (5) holds. (5) can be rewritten as

$$\mu \geq \alpha C(\gamma) + (1-\gamma)\theta,\ \alpha = \frac{1}{2} + \frac{1}{2}\frac{\gamma}{C(\gamma)} \geq 1. \tag{16}$$

Comparing the criterion for the Benthamite personal welfare maximum in (15) with the actual criterion in (16) we can see that the resolution is broken more often than welfare maximal since the right-hand side of (16) is larger for any given $\gamma < 1$. Especially, this is also true for $\gamma = \widehat{\gamma}$ since $\widehat{\gamma} < 2\theta - 1 < 1$. ■

Let us now for comparative reasons consider the welfare implications of the availability of external self-binding devices for S_p. The use of an external device by S_p restricts the action space of S_e to the singleton $\rho \in \{0\}$. S_p therefore excludes the possibility of a breach of resolution by S_e when using this option and will thus combine it with the strictest possible resolution of abstinence.

As we have seen, in a world without the availability of external self-binding devices S_p will choose a concessional level, i.e. form a resolution such that its expected utility is maximized. We called this concessional level $\widehat{\gamma}$. If we allow for the possibility of external self-binding the action set of S_p gets richer because it gets an additional strategic option.

S_p will choose to use an external self-binding device instead of $\widehat{\gamma}$, if its utility from external self-binding (S) is larger, $u^p(S) > E(u)^p(\widehat{\gamma})$, implying

$$(1-\phi)u^p(A) > ((1-\widehat{\gamma})u^p(A) + \frac{(\widehat{\gamma} - C(\widehat{\gamma}))}{2}(u^p(A) + u^e(A)))(1 - F(\widetilde{\mu}(\widehat{\gamma}|\theta)))$$

or

$$\widehat{\gamma} - \frac{\widehat{\gamma} - C(\widehat{\gamma})}{2\theta} + (\frac{\widehat{\gamma} - C(\widehat{\gamma})}{2\theta} + (1-\gamma))F(\widetilde{\mu}(\widehat{\gamma}|\theta) > \phi. \tag{17}$$

ϕ is the cost of the external self-binding device. This could be psychological costs of embarrassment or a reputational loss due to getting a third party involved.[3] We assume that both selves symmetrically suffer from embarrassment. If the external self-binding device is comparatively cheap, i.e. (17) holds, S_p will choose to use this device. Otherwise, S_p will prefer to make a concession of $\widehat{\gamma}$ giving it an expected utility which is at least as high.

In case (15) holds, it will be personal welfare maximizing that S_e adheres to the resolution, $\rho = 0$. In this case the external self-binding alternative will be welfare superior if $u^T(S) > u^T(\rho = 0|\gamma)$ implying

$$(1 - \phi)(u^p(A) + u^e(A)) > (1 - C(\gamma))(u^p(A) + u^e(A))$$

or

$$C(\gamma) > \phi. \tag{18}$$

If (15) does not hold, it will be welfare maximizing that S_e breaks the resolution, $\rho = 1$. Then, the external self-binding alternative will be welfare superior if $u^T(S) > u^T(\rho = 1|\gamma)$ implying

$$(1 - \phi)(u^p(A) + u^e(A)) > (1 - \mu)(u^p(A) + u^e(A))$$

or

$$\mu > \phi. \tag{19}$$

Proposition 5. *In equilibrium the external self binding device is used more frequently than a welfare maximum within a Benthamite framework suggests.*

Proof. Actually, S_p uses an external self-binding device if (17) holds. In case (15) holds, by comparing (17) and (18) we can see that external self-binding devices are used too frequently from a Benthamite personal welfare maximizing perspective. This is true since the left-hand side of the welfare maximizing criterion in (18) is smaller than the one of the actual criterion in (17). To see this, note that $\frac{2\theta\gamma - \gamma + C(\gamma)}{2\theta} \geq C(\gamma)$, assuring that the right-hand side of (17) is clearly bigger than $C(\gamma)$. In case (15) does not hold, the left-hand side of (19) is even smaller since $C(\gamma) > \mu$ by assumption. Therefore, we can see that self-binding

[3] An example causing embarrassment could be to admit a lack of self-control by giving away car keys to prevent oneself from drinking (see Elster 2000, p. 66). Monetary costs are less common but also imaginable. Bloom (2008) quotes a suggestion to remove your internet cable and FedEx it to yourself to have one day of work without online distractions. FedEx charges a monetary price, of course.

devices are used even more frequently than in the first case which gets us further away from the Benthamite personal welfare maximizing extent of use. Note that since this is true for any concessional level and its corresponding resolution it is, especially, true for the concession that S_p will choose, $\gamma = \widehat{\gamma}$. ■

Summing up these results we can see that when using a Benthamite welfare function the person will never achieve her welfare maximum. In a world without external self-binding devices resolutions are broken too often by trend since S_e is not taking the external effect of its actions on S_p into consideration. Accordingly, in a world with external self-binding devices S_p is making use of its costly power instrument too often. From a benevolent Benthamite dictator's perspective this is a dilemma that the person with the conflicting preferences we assumed cannot escape by herself. From her perspective a prohibition would seem appropriate since it leaves each self with $u^i(A)$, therefore maximizing personal welfare.

Note that the Benthamite position implies a clear value judgment which is of course arbitrarily chosen. Without going into details we would like to stress that switching to another personal welfare criterion could change welfare implications dramatically. Using a more egalitarian personal welfare criterion like the Bernoulli-Nash welfare function would condemn unequal utility distributions between the two selves and thus condemn a consumption prohibition in case of a strong inner disbalance. This is ignored by the Benthamite criterion. Refusing any kind of welfare function and applying a personal Pareto criterion as minimum consensus, it would judge each result the person achieves by herself as Pareto efficient since any deviation from this status quo would lead to a worsening of the situation of one self.

6. CONCLUSION

There is an emerging literature in behavioral economics trying to capture analytically the observation that human behavior often contradicts to neoclassical assumptions. We argue that giving up the assumption of the permanent identity between an acting person and a single economic agent can account for observed data without abandoning the neoclassical method.

This paper formulates a basic framework adapting a taxation model to explain the formation of resolutions. We assume the sequential existence of two agents in a person, being induced by different states and being in conflict with each other. This is a conflict that they are resolving in the intertemporal dimension. As there is no direct communication between the agents and though no kind of direct power contest, the only way to moderate consumption of a future agent is to form a more or less generous resolution or to use an external self-binding device. The final choice of an adequate device results from the utility maximization approach and depends both on the environmental expectations of the Stackelberg leader and the specific device costs. We have found unique equilibrium solutions allowing concrete empirical tests.

Without the suggestion of a concrete experimental design to test the predictions for certain parameter constellations we would like to stress the crucial point of state-dependency. One needs two different frames inducing conflicting individual choices to activate one of the agents in each frame. Participants would have to experience their conflict several times before their inner disbalance is measured as it is crucial that they are well aware of it. The less artificial and the closer to real life experience the situation is the better it works.

Our discussion about the welfare implications of our analysis can be seen as a postulation to make the implicit value judgments of paternalistic policies explicit. These policies mostly promote the interests of a long-term interested self which could be compared to the planning self in our model. As pointed out by other authors before, this is at least a debatable value judgment. Primarily, the presented framework wants to sensitize policy makers to be self-reflexive and transparent concerning their underlying norms. Gul and Pesendorfer (2008) criticize heavily the use of multiple selves models in the context of hyperbolic discounting frameworks for the arbitrariness of the welfare criteria used by the respective authors. Even though we share Gul and Pesendorfer's objection we do not think that this is a problem of the multiple selves notion itself but of the unreflected way in which welfare criteria are applied. The notion of multiple selves does, on the contrary, suggest an emancipation of selves as default. Finally, it is the general question of human rationality which hovers above our model. Once taking on the Heraclitean perspective an unquestionable best interest of a person no longer exists.

7. APPENDIX

The expected utility of S_p is

$$E(u)^p(\gamma|\theta) = ((1-\gamma)u^p(A) + \frac{(\gamma - C(\gamma))}{2}(u^p(A) + u^e(A)))\int_{\widetilde{\mu}(\gamma|\theta)}^{1} f(\mu)d\mu =$$

$$((1-\gamma)u^p(A) + \frac{(\gamma - C(\gamma))}{2}(u^p(A) + u^e(A)))(1 - F(\widetilde{\mu}(\gamma|\theta))).$$

Optimization calculus over γ implies

$$\frac{d}{d\gamma}E(u)^p(\gamma|\theta) = 0,$$

which means

$$\frac{d}{d\gamma}\ (1-\gamma)u^p(A)(1 - F(\widetilde{\mu}(\gamma|\theta)))$$

$$+\frac{d}{d\gamma}\ \frac{\gamma}{2}(u^p(A) + u^e(A))(1 - F(\widetilde{\mu}(\gamma|\theta)))$$

$$-\frac{d}{d\gamma}\ \frac{C(\gamma)}{2}(u^p(A)+u^e(A))(1-F(\widetilde{\mu}(\gamma|\theta)))=0,$$

or

$$-u^p(A)(1-F(\widetilde{\mu}(\gamma|\theta)))-(1-\gamma)u^p(A)f(\widetilde{\mu}(\gamma|\theta))\frac{d\widetilde{\mu}(\gamma|\theta)}{d\gamma}$$

$$+\frac{1}{2}(u^p(A)+u^e(A))(1-F(\widetilde{\mu}(\gamma|\theta)))-\frac{\gamma}{2}(u^p(A)+u^e(A))f(\widetilde{\mu}(\gamma|\theta))\frac{d\widetilde{\mu}(\gamma|\theta)}{d\gamma}$$

$$-\frac{\gamma}{2}(u^p(A)+u^e(A))(1-F(\widetilde{\mu}(\gamma|\theta)))+\frac{C(\gamma)}{2}(u^p(A)+u^e(A))f(\widetilde{\mu}(\gamma|\theta))\frac{d\widetilde{\mu}(\gamma|\theta)}{d\gamma}=0.$$

Dividing the equation by $(u^p(A)+u^e(A))$ and rearranging it gives

$$\frac{f(\widetilde{\mu}(\widehat{\gamma}|\theta))}{1-F(\widetilde{\mu}(\widehat{\gamma}|\theta))}=\frac{\theta+\frac{\widehat{\gamma}-1}{2}}{(\frac{C(\widehat{\gamma})-\widehat{\gamma}}{2}-(1-\widehat{\gamma})\theta)\frac{d\widetilde{\mu}(\widehat{\gamma}|\theta)}{d\gamma}}.$$

8. REFERENCES

Acemoglu, Daron and James Robinson (2001), **A Theory of Political Transitions.** The American Economic Review, vol. 91(4), pp. 938 - 963

Acemoglu, Daron and James Robinson (2006), **Economic Origins of Dictatorship and Democracy.** Cambridge University Press

Ainslie, George (2001), **Breakdown of Will.** Cambridge University Press

Bénabou, Roland and Jean Tirole (2004), **Willpower and Personal Rules.** Journal of Political Economy, vol. 112, no. 4, pp. 448 - 886

Bloom, Paul (2008), **First Person Plural.** The Atlantic, November

Brocas, Isabelle and Juan D. Carrillo (2008), **The Brain as a Hierarchical Organization.** The American Economic Review, vol. 98(4), pp. 1312 - 1346

Elster, Jon (2000), **Ulysses Unbound.** Cambridge University Press

Festinger, Leon (1957), **A Theory of Cognitive Dissonance.** Stanford University Press

Fudenberg, Drew and David Levine (2006), **A Dual-Self Model of Impulse Control.** The American Economic Review, vol. 96(5), pp. 1449 - 1476

Gul, Faruk and Wolfgang Pesendorfer (2008), **The Case for Mindless Economics.** Published in: Caplin, Andrew and Andrew Schotter (2008), The Foundations of Positive and Normative Economics. A Handbook. Oxford University Press, pp. 3 - 39

Güth, Werner (1991), **Game Theory's Basic Question: Who Is a Player?** Journal of Theoretical Politics, vol. 3(4), pp. 403 - 435

Karni, Edi (1993), **A Definition of Subjective Probabilities with State-Dependent Preferences.** Econometrica, vol. 61(1), pp. 187 - 198

Laibson, David (1997), **Golden Eggs and Hyperbolic Discounting.** Quarterly Journal of Economics, vol. CXII(2), pp. 443 - 477

Moldoveanu, Mihnea and Howard Stevenson (2001), **The self as a problem: the intra-personal coordination of conflicting desires.** Journal of Socio-Economics, vol. 30, pp. 295 - 330

O'Donoghue, Ted and George Loewenstein (2005), **Animal Spirits: Affective and Deliberative Processes in Economic Behavior.** Working Papers, Cornell University, Center for Analytic Economics

O'Donoghue, Ted and Matthew Rabin (1999), **Doing It Now or Later.** The American Economic Review, vol. 89(1), pp. 103 - 124

O'Donoghue, Ted and Matthew Rabin (2001), **Choice and Procrastination.** Quarterly Journal of Economics, vol. 116(1), pp. 121 - 160

Schelling, Thomas (1984), **Self-Command in Practice, in Policy, and in a Theory of Rational Choice.** The American Economic Review, vol. 74(2), pp. 1 - 14

Smith, Adam (1759), **The Theory of Moral Sentiments.** London: A. Millar

Strotz, Robert (1956), **Myopia and Inconsistency in Dynamic Utility Maximization.** The Review of Economic Studies, vol. 23(3), pp. 165 - 180

Thaler, Richard and Hersh Shefrin (1981), **An Economic Theory of Self-Control.** Journal of Political Economy, vol. 89(2), pp. 392 - 406